Open Journal of Nursing, 2014, 4, 774-783
Published Online October 2014 in SciRes. http://www.scirp.org/journal/ojn
http://dx.doi.org/10.4236/ojn.2014.411083

A Comparison of the NYHA Classification and the Duke Treadmill Score in Patients with Cardiovascular Disease

Sandra L. Carroll[1], Karen Harkness[2], Michael Hugh McGillion[1]

[1]McMaster University, Hamilton, Canada
[2]Hamilton Health Sciences, Hamilton, Canada
Email: carroll@mcmaster.ca

Received 19 August 2014; revised 20 September 2014; accepted 15 October 2014

Abstract

Aim: A discussion of the measurement of functional status in cardiovascular research. Background: Selection of appropriate outcome measures and instrumentation is vital for nurse researchers to ensure that outcomes align conceptually and are measured using reliable and valid tools. Functional status is a well-known outcome in cardiovascular research—however attention to underlying conceptual differences that may influence choices of whether to include subjective or objective measures is needed. Design: This work is a discussion paper: seminal research reporting the development, validation, and reliability testing of the NYHA classification system and the Duke Treadmill Score. Implications for nursing: Conceptually clarity and comprehensive appraisal of the reliability and validity of outcome measures in nursing research are essential to ensure high level of quality outcomes that will benefit patients.

Keywords

Nursing Practice, Function, Measurement, Cardiovascular Disease, Framework

1. Introduction

In nursing research, the selection of instruments and outcome measures to assess the effectiveness of health-related interventions requires careful consideration. To avoid mismatches, concepts of interest should link appropriately to the selected measures. In addition, the reliability and validity need to be clearly established. When the population of interest includes patients with cardiovascular disease (CVD), measurement and evaluation of functional status have been widespread in nursing research across health care settings and countries. Frequently, the New York Heart Association (NYHA) classification system, a subjective measure for assessing functional

How to cite this paper: Carroll, S.L., Harkness, K. and McGillion, M.H. (2014) A Comparison of the NYHA Classification and the Duke Treadmill Score in Patients with Cardiovascular Disease. *Open Journal of Nursing*, **4**, 774-783.
http://dx.doi.org/10.4236/ojn.2014.411083

status in patients with CVD, is used [1]. In the literature, the NYHA system is reported as a serial measurement of functional status, an inclusion criterion, as well as an outcome measure [2]. In cardiovascular research, when associations between the NYHA and health-related outcomes are proposed, the validity and reliability of this functional classification system require careful appraisal.

Alternatively, exercise stress testing offers an objective and reproducible measure of functional capacity [3]. The Duke Treadmill Score (DTS), for example, applies results from standard exercise stress testing to categorize patients with suspected or established CVD as high, moderate, or low risk for negative health outcomes [4] [5]. The DTS and the NYHA system provide options for objective and subjective measurement of functional status, respectively, cardiovascular nurse researchers. In what follows in this paper, we will apply a conceptual framework to examine the utility of two common measures of functional status.

2. Conceptual Framework

Maintaining and improving functional status represent meaningful outcomes to patients with cardiovascular related illnesses; however, ambiguity in the literature regarding both the measurement and conceptual interpretation of functional status pose challenges to employing appropriate instrumentation [6]-[8]. A useful vehicle to enhance conceptual clarity is the application of a framework, such as Leidy's conceptual framework [9]. This framework proposes four discrete dimensions of functional status: capacity, performance, reserve, and capacity utilization. Functional capacity represents the maximum potential of which an individual is capable. Physically, capacity reflects a person's trained state through ventilatory and cardiac capacity. Evaluation of capacity may include methods such treadmill testing, pulmonary function, and blood analysis. Functional performance represents the day-to-day activities a person *chooses* to undertake, within the limits of functional capacity. Motivation to perform activities of daily living influences a person's functional performance. As a person approaches his or her functional capacity, additional exertion is required to obtain a higher level of functional performance. As such, application of Leidy's [9] framework can provide researchers with clear direction on the measurement of functional status in cardiovascular patients as well as guidance in the development of appropriate interventions to impact functional status. In what follows, we will discuss two cardiovascular measures applying two of the definitions from Leidy's framework—capacity and performance.

3. NYHA Development

Developed by the Criteria Committee in 1964 with the purpose of creating a summary statement to describe the evaluation of cardiovascular status in patients, this classification system has since undergone two revisions [1] [10]. The system is based entirely on patients' subjective symptoms. The NYHA assigns functional Classes I - IV based on physical limitations caused by cardiac symptoms. The Classes describe progressive limitations in physical activity (Leidy's functional performance) ranging from minimal limitation (Class I) to an inability to continue any physical activity without symptoms (Class IV) [1]. The classification was expanded to include objective assessment classes (A - D) to complement functional capacity and to provide a more complete and "realistic" appraisal of cardiac status. The range of classes defines Class A patients as no objective evidence of CVD, Class B as moderately severe CVD, and up to Class D as evidence of severe CVD. The Committee provided no guidance or clear definition on the criteria used to determine these and is therefore a subjective process.

When the NYHA classification system was developed, it was not intended for use as a measurement instrument. Its purpose was to encourage physicians to provide the best therapeutic approach to care [1] [10]. Despite its reliance on the subjective interpretation of patient's symptoms, the functional classes are often erroneously applied during classification of ventricular dysfunction and congestive heart failure [1]. Because of the informal method of development, the system was not subjected to the traditional psychometric testing required of other tools used to measure health status [11] [12]. Therefore, assessment of the validity and reliability of the NYHA poses a challenge for researchers.

4. Validity & Reliability

Assessing validity can go beyond the psychometric properties of an instrument to examine *what* is being measured and the *relationship* of a variable to an underlying cause [12]. A valid instrument should precisely measure the intended concept and have low to moderate correlations with related concepts. Thus, if the concept of func-

tional status were associated with functional capacity and performance, as suggested by Leidy [9], a moderate correlation between measures purporting to assess these concepts would be expected [12]. Moreover, attention to the intended purpose of the NYHA in research studies must be given to determine the type of validity to appraise. To this end, the concurrent validity (a form of criterion validity) of the NYHA has been supported [13]-[15].

In three studies, NYHA functional class was validated against criterion measures of functional capacity in patients with heart failure [14] [15] and symptomatic or asymptomatic patients undergoing exercise stress testing [13]. Smith *et al*. [15] performed exercise testing using a defined bicycle ergometer protocol including direct measurement of peak oxygen uptake (VO_2). Goldman *et al*. [13] performed treadmill exercise testing using the Bruce protocol [16]. All three studies observed poor agreement between NYHA functional class and either peak VO_2 or exercise treadmill performance. Goldman *et al*. [13] reported 51% agreement between exercise performance and NYHA classes (weighted kappa 0.33). Similarly, Smith *et al*. [15] found inconsistency between NYHA classifications and peak VO_2 results within each NYHA class. Although there was a trend toward lower maximal VO_2 among Class III NYHA patients, Class I - II patients also reported low maximal VO_2 during exercise testing using a bicycle ergometer [15]. Lastly, Rostagno *et al*. [14] found only 41.7% agreement between peak VO_2 and NYHA classes.

Convergent validity, a form of construct validity, was assessed when the NYHA was compared with level of performance in heart failure patients using the six-minute walk test (6 MWT). In 33% of patients, categorization according to NYHA class agreed with distance walked in the 6 MWT [14]. Concordance was highest among patients with the poorest NYHA class, that is, those walking the shortest distances. Other investigations also support the inverse association between NYHA class and the 6 MWT in Class IV NYHA patients [17] [18]. Thus, the moderate associations between the NYHA and the 6 MWT suggest that these two measures assess similar theoretical constructs of function, namely Leidy's [9] functional performance.

Certain properties in the NYHA system, such as interpretability (classes) and scaling warrant further discussion of validity. Regarding interpretability, the most ambiguous language found in the NYHA is the term "ordinary physical activity" [1]. The meaning of "ordinary", followed by a subjective judgment on the degree of compromise, can vary among patients, among assessors, or over time as health conditions fluctuate [19]. In patients with CVD, voluntary limitation in activity because of poor cardiovascular health or other co morbid conditions might lead to diminished symptoms from lack of exertion. As a result, when questioned, patients may not report symptoms because "ordinary physical activity" is sedentary. Raphael *et al*. [2] examined the accuracy of responses to the most commonly asked question in patients with heart failure: "How far can you walk?" The study reported that the median estimation made by patients for a 50-metre distance was 73 metres (range 9 to 274 metres)—an overestimation of 46%. Attempts to improve the accuracy in NYHA class assignment have been made by developing adjunct questionnaires aimed at increasing precision [2] [20].

The NYHA uses an ordinal scale to reflect the increasing severity of cardiovascular symptoms. The limitation imposed by only four categories becomes apparent when discrimination between NYHA Class II and III is required. Determining the difference between "slight" and "marked" limitation of physical activity requires additional interpretation of physical activity during assessment [2] [10]. In Raphael *et al*.'s [2] survey of 50 cardiologists, 67% reported using a patient's ability to walk up a flight of stairs as the discriminatory question to determine a classification of II or III. The adjunct questionnaires developed by Goldman *et al*. [13] and Kubo *et al*. [20], aimed at increasing the accuracy of NYHA class assignment, were shown to be valid when used by physicians and non-physicians. The continued application of these adjunct questionnaires in research is not known.

Discriminating between the four NYHA classes becomes a primary concern when the NYHA is used as a criterion for inclusion in research. For example, when a study purports to include only NYHA Class III, the ability to accurately assign NYHA class can ultimately influence the research findings. Raphael *et al*. [2] examined the use of NYHA class in 200 clinical trials, finding that 80 of these used the NYHA as an inclusion criterion. Depending on the desired research outcome, systematic misclassification of NYHA class could lead to biased results. In addition, 99 of these trials used NYHA class as both an inclusion criterion and an outcome measure. However, evidence of guidelines or standardization applied to determine NYHA class is rarely considered or reported [2].

If NYHA classes are to be used as a measure in nursing research, then the responsiveness of the NYHA to change must be established. The terms "sensitivity" and "responsiveness" are often used interchangeably [12]. Streiner and Norman [12] consider sensitivity and responsiveness to be a component of the validity of a measure

rather than a third variable to include with reliability and validity. Sensitivity to change is "the ability of an instrument to measure change in a state regardless of whether it is relevant or meaningful to the decision maker" [21], whereas responsiveness is "the ability of an instrument to measure a meaningful or clinically important change" [21]. When one considers that the NYHA was not developed to be used as a measurement tool, it is surprising to find its widespread use as an outcome variable in clinical trials. Questions regarding sensitivity remain: Are the four NYHA classes sensitive enough to discrete changes within the same patient along the continuum of CVD? Is the NYHA responsive enough to measure change over time? Demers *et al*. [18] indirectly established the responsiveness of the NYHA among patients with chronic heart failure. The trial was powered to evaluate the reliability of the 6MWT as a measure of functional performance in patients with heart failure randomized to medical intervention or placebo. The study observed small changes in the NYHA demonstrating responsiveness to change in the sample; although small, these changes were purported to have clinical importance.

Few studies have directly evaluated the reliability of the NYHA classification system [2] [13] [22], thus, its inclusion as an outcome measure in clinical trials is concerning. Appropriate NYHA classification relies on the consistent interpretation of the rater. Test-retest reliability methods may not be reliable in CVD patients because of the shifting nature of symptoms that may occur between test comparisons.

Goldman *et al*. [13] used two physicians to estimate NYHA functional class in 75 patients on the same day without chronic heart failure, reporting an interrater reliability of 56% (weighted kappa = 0.41). In a second study, two cardiologists assessed the same 50 chronic heart failure patients on the same day in random order, observing 54% agreement in NYHA classes [2]. In a third study, two physicians assigned NYHA class to 56 patients [22] with stable angina within the same hour, resulting in the highest reported agreement of 75%. Among these studies, disagreement by more than one functional class was low and, for the most part, was concentrated on determining the discrete differences between Classes II and III. Taken together, the reliability of the NYHA system is limited in the few trials that have measured it directly.

5. Strengths and Weaknesses

The greatest strength of the NYHA classification system is its utility as a functional classification system that is meaningful across healthcare professionals when reporting on the clinical status of patients. Despite this, the NYHA has weaknesses. First, assigning NYHA Class I or IV is not problematic in a clinical setting, reliably discriminating between NYHA classes II and III poses greater difficulty. Questionable reliability limits the application of the NYHA class system as a research tool, especially when neither the objective components of the NYHA nor a description of methods to obtain the NYHA is reported [2]. Second, there is poor agreement between criterion measures of functional capacity and the NYHA when assessed at the same point in time, highlighting the limitation of the NYHA as a measure of functional capacity in patients with CVD [2] [13] [15]. Third, there is limited evidence of the NYHA's responsiveness to interventions; thus, its use as an outcome measure in trials re-evaluated.

In nursing research, utilization of the NYHA classification system as an outcome measure to assess changes in functional capacity or performance should be cautioned, for two reasons: the NYHA measures functional status, and there is limited evidence of the NYHA's responsiveness. Most importantly, without evidence of reliability, researchers cannot be certain that the NYHA is able to consistently assess functional status on the same patient either by different observers or on different occasions.

6. Duke Treadmill Score Development

The DTS was developed as a simple quantitative tool to assist in the stratification of prognosis and risk of future coronary events in patients [23]. The developers had three goals: to focus on prognosis rather than diagnosis, to create a single index based on independent prognostic data from a treadmill test, and to keep it simple and user friendly [23]. The treadmill score was originally derived from in patients referred for treadmill testing and angiography [4]. Variables obtained from treadmill testing were ranked according to prognostic importance in predicting cardiovascular death [4]. The score was derived from the three highest ranked variables: the amount of ST depression on an electrocardiogram, exercise capacity, and the presence or absence of angina [24]. The DTS incorporates the estimated maximal metabolic equivalent (MET) level, which is a surrogate measure of VO_2 (exercise capacity) attained by exercise testing using the Bruce treadmill protocol [24] [25]. Specific cutoffs for achievable MET levels have been associated with survival in CVD patients [26]. The advantage of the Bruce

protocol is that it does not require costly gas exchange to measure VO_2 directly but estimates it based on standardized values. The Bruce protocol reliably predicts the peak VO_2 (consumption) in most patients using the maximum duration of exercise (in minutes) [24]. ST depression is measured and the treadmill angina index coded from 0 to 2 [3]. The DTS incorporates these three variables to calculate a prognostic risk score. These variables are entered into a simple equation and calculated as follows [4]:

DTS = exercise time − (5 × ST deviation) − (4 × exercise angina)

Once a score is calculated, patients are stratified into three cardiovascular-related mortality risk categories: high risk (≤−11), moderate risk (−10 to +4), and low risk (≥5).

7. Validity and Reliability

In the seminal work by Mark *et al*. [4], the performance of the DTS was validated during development using a test sample from a study population of 2842 patients by splitting the sample into two groups, exercise training and test sample (control). Kaplan-Meier estimates of survival rates in both groups were compared using the DTS, observing similar rates in each sample and thus reproducible prognostic stratification and supporting the reproducibility of the DTS.

The original validation sample consisted of inpatients who were preselected to undergo angiography; therefore, it was unknown if the DTS would perform equally as well in an outpatient population with unknown disease status. To address this, Mark *et al*. [5] undertook a second validation study using the DTS in an unselected sample (n = 613) referred for non-invasive assessment of suspected CVD. The study aimed to examine two aspects of predictive accuracy of the DTS: (1) its reliability in agreement with observed prevalence of cardiovascular death and (2) how well the score discriminated between those who survived and those who did not. The DTS demonstrated good predictive accuracy in four-year survival rates for patients categorized as low, moderate, and high risk and confirmed that the DTS could reproduce prognostic stratification in both inpatient and outpatient populations [5]. These findings are important for nurse researchers when considering application of the DTS as a research tool, particularly the ability to identify patients at low risk for future cardiac events for inclusion in exercise-based intervention and secondary prevention trials. Moreover, identification of patients at low risk for future events using the DTS as a prognostic risk stratification measure after exercise testing may reduce the necessity of costly invasive procedures in patients who are unlikely to experience a negative cardiovascular outcome [3] [23].

Although the DTS was demonstrated to be valid among males with or without evidence of CVD, validation among other subgroups of patients is imperative for nurse researchers to ensure that the DTS can stratify among differences in risk in particular samples of patients. The DTS should demonstrate reproducibility among different groups of patients to consider including the tool in different populations [27]. To address this, validation studies to assess the predictive accuracy and reproducibility of the DTS were conducted in subgroups of women and elderly patients [28]-[30]. Two studies explored the predictive accuracy and reproducibility of the DTS in women. Alexander *et al*. [28] evaluated symptomatic women (n = 976) and men (n = 2249) who underwent a treadmill test and cardiac catheterization The two-year mortality risk for women and men was 1.9% and 4.9%, respectively. Among the low, moderate, and high DTS risk categories, the mortality rate for women was 1%, 2.2%, and 3.6%. For men, the mortality rates by DTS category were 1.7%, 5.8%, and 16.6%, respectively. The prevalence of CVD assessed by cardiac catheterization was 32% for women and 72% for men ($p < 0.001$). Overall, the DTS was found to have equivalent predictive abilities in men and women.

Gulati *et al*. [29] followed a cohort of 5636 asymptomatic women from 1992 to 2000 to examine the prognostic performance of the DTS and exercise capacity in relation to all-cause mortality. In this cohort, the mean DTS score was 8 ± 4 (low risk). Survival analysis revealed that hazard ratios for death were 2.2 times greater among women with a DTS < 5 compared with women with a DTS ≥ 5 (95% CI: 1.6 - 3.1). Exercise capacity and DTS were both independent predictors of mortality. For each unit of increase in the DTS (improved score), the hazard ratio decreased by 9%. The effect was more powerful for exercise capacity alone; for each one MET increase in exercise performance, the hazard ratio decreased by 17%. Therefore, the predictive value of the DTS was due primarily to the exercise capacity component of the score. Functional capacity in this cohort influenced all-cause mortality.

Finally, Kwok *et al*. [30] hypothesized that the DTS would not discriminate risk as well in the elderly (>75

years) compared with non-elderly controls (<75 years) over a median period of 6.5 years. In the elderly group, the proportions of patients with low, moderate, and high DTS risk categories were 26%, 68%, and 6%, respectively (n = 247), but these scores did not predict the cardiovascular-related outcomes as well in the elderly groups compared to the non-elderly group (n = 2304). The DTS in this relatively small sample did not perform as well compared with controls. However, the elderly had a shorter exercise time, making it more difficult to achieve greater time. The mean DTS score was also significantly lower in the elderly group at 2.3 (range −3.2 to 5.0) compared with the controls at 5.5 (range 0.5 to 8.4). Exercise testing in elderly patients using a treadmill may be limited due to inability to achieve exercise times because of overall deconditioning.

Although developed as a prognostic tool, the DTS has been validated as a diagnostic tool [31] [32]. Lipinski *et al*. [31] compared the probability estimates of CAD severity and death made by physicians with estimates from four different treadmill scores in patients with chest pain (n = 599). Receiver operating curves (ROC) were used to analyze the predictive accuracy of the treadmill and physician estimates of disease severity. Here, the DTS performed as well or better than physician judgment in predicting significant CAD. Shaw *et al*. [32] also observed that the DTS provided accurate diagnostic information to predict significant CAD (≥75% stenosis); the area under the ROC was 0.76 in a sample of 467 patients. Both studies used angiograms as the criterion measure for comparison. The findings demonstrate that the DTS may accurately predict the presence of CAD and could be considered when nurses are planning research for primary prevention interventions. Application of the DTS as a diagnostic tool in secondary prevention studies is unwarranted and should be considered only when risk stratification of patients is advised.

The reliability of the DTS as measured by interrater, intrarater, or test-retest methods was not reported in these validation studies. Although considered an objective tool, the DTS *is* susceptible to random or systematic errors in measurement during evaluation of patient symptoms. This error could occur within any of the three variables that make up the DTS. The Bruce treadmill protocol, which is prescriptive in its six stages, timing, and measurement, also includes the Borg scale (a subjective instrument to measure the level of exertion and breathlessness during exercise testing) to assess limiting symptoms during testing [33]. The treadmill itself requires periodic calibration to maintain accuracy and is less accurate when patients hold the handrails during testing, which, in turn, influences exercise capacity and electrocardiogram lead positioning [33]. Incorrect positioning of chest leads for measuring ST changes could lead to discrepancies and influence calculation of the total DTS. The sensitivity of exercise electrocardiography in measuring true ST changes in patients is embedded within the DTS and, as such, requires its own reliability and validity as part of the DTS [34]. Because the DTS is a composite tool, all three variables must possess good validity and reliability for the score to be accurate.

8. Strengths and Weaknesses

As a prognostic risk tool, the body of evidence suggests that the DTS effectively and consistently stratifies both men and women, with or without chest pain symptoms, into low-, moderate-, and high-risk groups for future negative cardiovascular outcomes [4] [5] [28] [29]. Moreover, the DTS performs well in both inpatient and outpatient cardiovascular populations of patients with CVD [4] [5]. However, the DTS is not as precise when stratifying elderly patients (>75 years) [30]. As a diagnostic tool, the DTS demonstrated predictive accuracy for determining the presence of significant CAD [31] [32]. Overall, the DTS is a valid tool when used for risk stratification purposes and is recommended in practice guidelines for exercise testing [35] [36]. In addition, the DTS can indirectly provide an indication of functional capacity in patients through exercise times achieved using the Bruce protocol and corresponding achieved METS, without the added costs and resources required to assess VO_2 (the gold standard).

As a research tool in cardiovascular patients, the DTS can offer a simple, inexpensive risk stratification to obtain baseline risk scores in patients, prior to inclusion into secondary prevention trials aimed at improving functional capacity or improving health-related quality of life (HRQoL).

The DTS is not without limitations. The responsiveness of the DTS through serial testing was not reported directly in the literature we reviewed. Studies using the DTS focused primarily on a single measurement to determine patient limitations imposed by CAD and the tool's prognostic significance. The magnitude of change that is clinically meaningful in response to interventions, as measured by the DTS, is unclear. Nurses may choose a reduction in a risk category from moderate to low as a possible research outcome. However, before using the DTS as an outcome variable to measure risk reduction, additional evidence demonstrating its respon-

siveness to change is required. Moreover, the DTS's test-retest reliability should be considered, and assessed. The American Heart Association recommends that exercise testing not vary more than 10% on repeated testing [37]. Therefore, in nursing research, measures to ensure the DTS's reliability must be defined and adhered to. These measures could include creation of standardized operation manuals, training of research staff, and administering the exercise tests at the same time of day [33].

9. Conclusions and Implications for Nursing Research

The development of research proposals designed to answer questions that are important to nursing, careful attention to instrumentation and properties of measurement tools used to gather outcome data can avoid unanswered questions about study findings. Developing a research question that aims to explore "function" as an outcome necessitates an examination of available methods developed to reliably measure the construct [38]. Moreover, appraisal of psychometric evidence to support the reliability and validity of such measures is required. Finally, to guide nurse researchers during these decisions, the application of a conceptual framework such as Leidy's [9] strengthens the internal validity of the research. In this paper, we discussed two measures of function, as such; measurement of function in terms of "capacity" and "performance" using two well known measures provided a concrete example of the complexities that researchers may face. Assuming that "function" is measured reliably using the NYHA classification system as an outcome variable may lead to difficulties during interpretation of the study findings. The NYHA and the DTS propose to measure components of physical function. Upon closer examination, each tool has limitations related to how and when the concept of function is adequately represented. The NYHA remains a simple, subjective classification system for patients with CVD. Because of its current widespread adoption, it will likely continue as a measure to classify patients in clinical settings and as a research criterion. Moderate correlations observed with associated measures of functional capacity and performance may suggest that a more effective use of NYHA classes could be as an "anchor" to make comparisons with HRQoL [39]. Used in this way, the NYHA class measured at baseline could provide an anchor to HRQoL or functional measure outcome. However, nurses considering inclusion of the NYHA should clarify the dimension of function they are addressing in the research question. Functional status is not just capacity or performance, and the NYHA was not developed to adequately represent these specific dimensions. Nurse researchers who choose the NYHA as an inclusion criterion should, 1) include a description of how the classes are assigned by raters, 2) include patients as self raters [40], 3) provide the objective components of the NYHA, 4) include evidence of interrater or intrarater reliability, and 5) administer training to raters.

The language and interpretability of the NYHA remain ambiguous, making patients subject to potential biases in responses depending on the situation, state of health, and environment. The NYHA is subjective, and patients often describe symptoms while sitting in a chair; these symptoms are then interpreted by a rater, who assigns them an NYHA class. The NYHA does not take into account physical limitations that interfere with the ability of a person to perform "normal activities". In addition, if patients rate current symptom status against a previous point in time, they may shift responses in light of their current state of health [12] [40]. Living with long-term, multiple, chronic health conditions can result in patients' adjustment to illness and a redefining of what "normal" is for them. In cardiovascular research, careful attention to choosing a measurement tool that captures the dimensions influenced by interventions will result in research outcomes that are more meaningful to both patients and healthcare providers.

The DTS as a research tool offers nurse researchers an alternative or adjunct to subjective measures of function and demonstrates strong consistent measurement properties. As a modifiable risk factor, poor exercise capacity identified through exercise testing may be amenable to exercise-based interventions. However, not all patients with CVD will be suitable or willing to perform repeated exercise tests on a treadmill as part of a research protocol. In particular, patients with heart rhythm disorders, pacemakers, or implantable defibrillators, who have conduction abnormalities (complete left bundle branch block, paced rhythm), can greatly affect the diagnostic performance of the exercise test [36]. Depending on the patient population, alternative measures of function might be required. Furthermore, obtaining patients' perception of their function is important, and perhaps blending the two types of instruments could add value in nursing research. Keeping Leidy's framework in the forefront during study design decisions, researchers must ask themselves if the aim of the intervention is to improve "capacity" or "performance" in the context of the NYHA or DTS. However, researchers are not limited ot the NYHA and DTS, several tools are available to consider [38].

Nurse researchers may consider a number of additional factors when choosing measurement instruments. These include but are not limited to instrument or testing cost, availability for public use, ease of use (scale, number of items), patient and/or resource burden, and method of administration (telephone, mail, in person). Inclusion of a conceptual framework to guide the researcher during selection of measures is imperative.

In summary, thoughtful selection of outcome measures and instrumentation that can effectively bring to light meaningful change from nurse-led interventions will benefit patients with CVD.

Acknowledgements

Dr. S.L. Carroll is a recipient of a Research Early Career Award from Hamilton Health Sciences Foundation. Dr. M.H. McGillion holds the Heart & Stroke Foundation/Michael G. DeGroote Endowed Chair in Cardiovascular Nursing Research at McMaster University.

Conflict of Interests

No conflict of interest has been declared by the authors.

References

[1] Criteria Committee of the New York Heart Association (1994) Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels. 9th Edition, Little Brown, Boston.

[2] Raphael, C., Briscoe, C., Davies, J., Ian, W.Z., Manisty, C., Sutton, R., *et al*. (2007) Limitations of the New York Heart Association Functional Classification System and Self-Reported Walking Distances in Chronic Heart Failure. *Heart*, **93**, 476-482. http://dx.doi.org/10.1136/hrt.2006.089656

[3] Froelicher, V., Shetler, K. and Ashley, E. (2002) Better Decisions through Science: Exercise Testing Scores. *Progress in Cardiovascular Diseases*, **44**, 395-414. http://dx.doi.org/10.1053/pcad.2002.122693

[4] Mark, D.B., Hlatky, M.A., Harrell Jr., F.E., Lee, K.L., Califf, R.M. and Pryor, D.B. (1987) Exercise Treadmill Score for Predicting Prognosis in Coronary Artery Disease. *Annals of Internal Medicine*, **106**, 793-800. http://dx.doi.org/10.7326/0003-4819-106-6-793

[5] Mark, D.B., Shaw, L., Harrell Jr., F.E., Hlatky, M.A., Lee, K.L., Bengtson, J.R., *et al*. (1991).Prognostic Value of a Treadmill Exercise Score in Outpatients with Suspected Coronary Artery Disease. *New England Journal of Medicine*, **325**, 849-853. http://dx.doi.org/10.1056/NEJM199109193251204

[6] Coyne, K.S. and Allen, J.K. (1998) Assessment of Functional Status in Patients with Cardiac Disease. *Heart & Lung*, **27**, 263-273. http://dx.doi.org/10.1016/S0147-9563(98)90038-3

[7] Miller-Davis, C., Marden, S. and Leidy, N.K. (2006) The New York Heart Association Classes and Functional Status: What Are We Really Measuring? *Heart & Lung*, **35**, 217-224. http://dx.doi.org/10.1016/j.hrtlng.2006.01.003

[8] Wang, T. (2004) Concept Analysis of Functional Status. *International Journal of Nursing Studies*, **41**, 457-462. http://dx.doi.org/10.1016/j.ijnurstu.2003.09.004

[9] Leidy, N.K. (1994) Functional Status and the Forward Progress of Merry-Go-Rounds: Toward a Coherent Analytical Framework. *Nursing Research*, **43**, 196-202. http://dx.doi.org/10.1097/00006199-199407000-00002

[10] Criteria Committee of the New York Heart Association, Harvey, R.M., Doyle, E.F., Ellis, K., Farber, S.J., Ferrier, M.I., *et al*. (1974) Major Changes Made by Criteria Committee of the New York Heart Association. *Circulation*, **49**, 390. http://dx.doi.org/10.1161/01.CIR.49.3.390

[11] Selzer, A. and Cohn, K. (1972) Functional Classification of Cardiac Disease: A Critique. *American Journal of Cardiology*, **30**, 306-308. http://dx.doi.org/10.1016/0002-9149(72)90079-3

[12] Streiner, D.L. and Norman, G.R. (2003) Health Measurement Scales: A Practical Guide to Their Development and Use. 3rd Edition, Oxford University Press, Oxford.

[13] Goldman, L., Hashimoto, B., Cook, E.F. and Loscalzo, A. (1981) Comparative Reproducibility and Validity of Systems for Assessing Cardiovascular Functional Class: Advantages of a New Specific Activity Scale. *Circulation*, **64**, 1227-1234. http://dx.doi.org/10.1161/01.CIR.64.6.1227

[14] Carlo, R., Giorgio, G., Marco, C., Vieri, B., Giuseppe, O. and Gastone Neri, S.G. (2000) Comparison of Different Methods of Functional Evaluation in Patients with Chronic Heart Failure. *European Journal of Heart Failure*, **2**, 273-280. http://dx.doi.org/10.1016/S1388-9842(00)00091-X

[15] Smith, R.F., Johnson, G., Ziesche, S., Bhat, G., Blankenship, K. and Cohn, J.N. (1993) Functional Capacity in Heart Failure. Comparison of Methods for Assessment and Their Relation to Other Indexes of Heart Failure. The V-HeFT VA Cooperative Studies Group. *Circulation*, **87**, VI88-VI93.

[16] Bruce, R.A. (1971) Exercise Testing of Patients with Coronary Heart Disease. Principles and Normal Standards for Evaluation. *Annals of Clinical Research*, **3**, 323-332.

[17] Bittner, V., Weiner, D.H., Yusuf, S., Rogers, W.J., McIntyre, K.M., Bangdiwala, S.I., *et al*. (1993) Prediction of Mortality and Morbidity with a 6-Minute Walk Test in Patients with Left Ventricular Dysfunction. *JAMA*, **270**, 1702-1707. http://dx.doi.org/10.1001/jama.1993.03510140062030

[18] Demers, C., McKelvie, R.S., Negassa, A. and Yusuf, S., for the RESOLVD Pilot Study Investigators (2001) Reliability, Validity, and Responsiveness of the Six-Minute Walk Test in Patients with Heart Failure. *American Heart Journal*, **142**, 698-703. http://dx.doi.org/10.1067/mhj.2001.118468

[19] Goldman, L., Cook, E.F., Mitchell, N., Flatley, M., Sherman, H. and Cohn, P.F. (1982) Pitfalls in the Serial Assessment of Cardiac Functional Status: How a Reduction in "Ordinary" Activity May Reduce the Apparent Degree of Cardiac Compromise and Give a Misleading Impression of Improvement. *Journal of Chronic Diseases*, **35**, 763-771. http://dx.doi.org/10.1016/0021-9681(82)90087-X

[20] Kubo, S.H., Schulman, S., Starling, R.C., Jessup, M., Wentworth, D. and Burkhoff, D. (2004) Development and Validation of a Patient Questionnaire to Determine New York Heart Association Classification. *Journal of Cardiac Failure*, **10**, 228-235. http://dx.doi.org/10.1016/j.cardfail.2003.10.005

[21] Liang, M.H. (2000) Longitudinal Construct Validity: Establishment of Clinical Meaning in Patient Evaluative Instruments. *Medical Care*, **38**, II84-II90. http://dx.doi.org/10.1097/00005650-200009002-00013

[22] Christensen, H.W., Haghfelt, T., Vach, W., Johansen, A. and Hoilund-Carlsen, P.F. (2006) Observer Reproducibility and Validity of Systems for Clinical Classification of Angina Pectoris: Comparison with Radionuclide Imaging and Coronary Angiography. *Clinical Physiology & Functional Imaging*, **26**, 26-31. http://dx.doi.org/10.1111/j.1475-097X.2005.00643.x

[23] Mark, D.B. (1994) An Overview of Risk Assessment in Coronary Artery Disease. *American Journal of Cardiology*, **73**, B19-B25. http://dx.doi.org/10.1016/0002-9149(94)90261-5

[24] Ashley, E.A., Myers, J. and Froelicher, V. (2000) Exercise Testing in Clinical Medicine. *Lancet*, **356**, 1592-1597. http://dx.doi.org/10.1016/S0140-6736(00)03138-X

[25] Franklin, B.A. (1995) Diagnostic and Functional Exercise Testing: Test Selection and Interpretation. *Journal of Cardiovascular Nursing*, **10**, 8-29. http://dx.doi.org/10.1097/00005082-199510000-00003

[26] Morris, C.K., Ueshima, K., Kawaguchi, T., Hideg, A. and Froelicher, V.F. (1991) The Prognostic Value of Exercise Capacity: A Review of the Literature. *American Heart Journal*, **122**, 1423-1431. http://dx.doi.org/10.1016/0002-8703(91)90586-7

[27] DiCenso, A. and Guyatt, G. (2005) Prognosis. In: DiCenso, A., Guyatt, G. and Ciliska, D., Eds., *Evidence Based Nursing*: *A Guide to Clinical Practice*, Elsevier Mosby, St. Louis, 108-119.

[28] Alexander, K.P., Shaw, L.J., DeLong, E.R., Mark, D.B. and Peterson, E.D. (1998) Value of Exercise Treadmill Testing in Women. *Journal of the American College of Cardiology*, **32**, 1657-1664. http://dx.doi.org/10.1016/S0735-1097(98)00451-3

[29] Gulati, M., Arnsdorf, M.F., Shaw, L.J., Pandey, D.K., Thisted, R.A., Lauderdale, D.S., *et al*. (2005) Prognostic Value of the Duke Treadmill Score in Asymptomatic Women. *American Journal of Cardiology*, **96**, 369-375. http://dx.doi.org/10.1016/j.amjcard.2005.03.078

[30] Kwok, J.M., Miller, T.D., Hodge, D.O. and Gibbons, R.J. (2002) Prognostic Value of the Duke Treadmill Score in the Elderly. *Journal of the American College of Cardiology*, **39**, 1475-1481. http://dx.doi.org/10.1016/S0735-1097(02)01769-2

[31] Lipinski, M., Froelicher, V., Atwood, E., Tseitlin, A., Franklin, B., Osterberg, L., Do, D. and Myers, J. (2002) Comparison of Treadmill Scores with Physician Estimates of Diagnosis and Prognosis in Patients with Coronary Artery Disease. *American Heart Journal*, **143**, 650-658. http://dx.doi.org/10.1067/mhj.2002.120967

[32] Shaw, L.J., Peterson, E.D., Shaw, L.K., Kesler, K.L., DeLong, E.R., Harrell Jr., F.E., *et al*. (1998) Use of a Prognostic Treadmill Score in Identifying Diagnostic Coronary Disease Subgroups. *Circulation*, **98**, 1622-1630. http://dx.doi.org/10.1161/01.CIR.98.16.1622

[33] American College of Sports Medicine (2006) ACSM's Resource Manual for Guidelines for Exercise Testing and Prescription. 7th Edition, Lippincott Williams & Wilkins, Baltimore.

[34] Hlatky, M.A., Pryor, D.B., Harrell Jr., F.E., Califf, R.M., Mark, D.B. and Rosati, R.A. (1984) Factors Affecting Sensitivity and Specificity of Exercise Electrocardiography. Multivariable Analysis. *American Journal of Medicine*, **77**, 64-71. http://dx.doi.org/10.1016/0002-9343(84)90437-6

[35] Balady, G.J., Williams, M.A., Ades, P.A., Bittner, V., Comoss, P., Foody, J.M., Franklin, B., Sanderson, B. and Southard, D. (2007) Core Components of Cardiac Rehabilitation/Secondary Prevention Programs: 2007 Update: A Scientific Statement from the American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the

Council on Clinical Cardiology; the Councils on Cardiovascular Nursing, Epidemiology and Prevention, and Nutrition, Physical Activity, and Metabolism; and the American Association of Cardiovascular and Pulmonary Rehabilitation. *Circulation*, **115**, 2675-2682. http://dx.doi.org/10.1161/CIRCULATIONAHA.106.180945

[36] Gibbons, R.J., Balady, G.J., Beasley, J.W., Bricker, J.T., Duvernoy, W.F., Froelicher, V.F., *et al.* (1997) ACC/AHA Guidelines for Exercise Testing: Executive Summary. A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing). *Circulation*, **96**, 345-354. http://dx.doi.org/10.1161/01.CIR.96.1.345

[37] Fleg, J.L., Pina, I.L., Balady, G.J., Chaitman, B.R., Fletcher, B., Lavie, C., *et al.* (2000) Assessment of Functional Capacity in Clinical and Research Applications: An Advisory from the Committee on Exercise, Rehabilitation, and Prevention, Council on Clinical Cardiology, American Heart Association. *Circulation*, **102**, 1591-1597. http://dx.doi.org/10.1161/01.CIR.102.13.1591

[38] Kocks, J.W.H., Asijee, G.M., Tsiligianni, I.G., Kerstjens, H.A.M. and van der Molen, T. (2011) Functional Status Measurement in COPD: A Review of Available Methods and Their Feasibility in Primary Care. *Primary Care Respiratory Journal*, **20**, 269-275. http://dx.doi.org/10.4104/pcrj.2011.00031

[39] Guyatt, G.H., Osoba, D., Wu, A.W., Wyrwich, K.W. and Norman, G.R., Clinical Significance Consensus Meeting Group (2002) Methods to Explain the Clinical Significance of Health Status Measures. *Mayo Clinic Proceedings*, **77**, 371-383. http://dx.doi.org/10.4065/77.4.371

[40] Holland, R., Stepien, K., Harvey, I. and Brooksby, I. (2010) Patients' Self-Assessed Functional Status in Heart Failure by New York Heart Association Class: A Prognostic Predictor of Hospitalizations, Quality of Life and Death. *Journal of Cardiac Failure*, **16**, 150-156. http://dx.doi.org/10.1016/j.cardfail.2009.08.010